(12) UK Patent Application (19) GB (11) 2 332 544 (13) A

(43) Date of A Publication 23.06.1999

(21) Application No 9827135.6

(22) Date of Filing 09.12.1998

(30) Priority Data
(31) 08995616 (32) 22.12.1997 (33) US

(51) INT CL[6]
G06F 17/30

(52) UK CL (Edition Q)
G4A AUXX

(56) Documents Cited
EP 0802492 A1 EP 0762297 A2 EP 0378848 A2
US 5404295 A

(58) Field of Search
UK CL (Edition Q) G4A AUDB AUXX
INT CL[6] G06F

(71) Applicant(s)
Ricoh Company Limited
(Incorporated in Japan)
No 3-6 Nakamagome 1-chome, Ohta-ku, Tokyo, Japan

(72) Inventor(s)
Jamey Graham
David G. Stork

(74) Agent and/or Address for Service
Marks & Clerk
57-60 Lincoln's Inn Fields, LONDON, WC2A 3LS,
United Kingdom

(54) Abstract Title
Automatic adaptive document help system

(57) In an automatic reading assistance system for electronic documents (502), an automatic annotator (508) finds concepts of interest (512) and keywords. The operation of the annotator is personalizable (518) for a particular user. The annotator is also capable of improving its performance over time by both automatic and manual feedback. Another available feature is a thumbnail image of all or part of a multi-page document wherein a currently displayed section of the document is highlighted in the thumbnail image. Movement of the highlighted area in the thumbnail image is then synchronized with scrolling through the document.

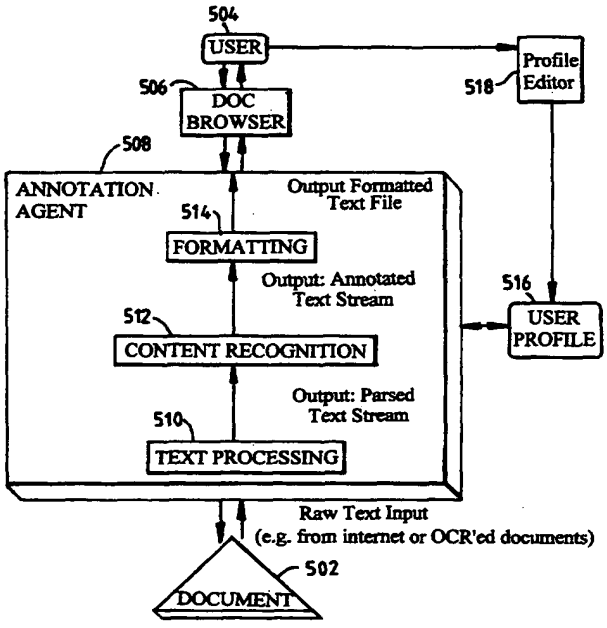


Fig. 5

GB 2 332 544 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

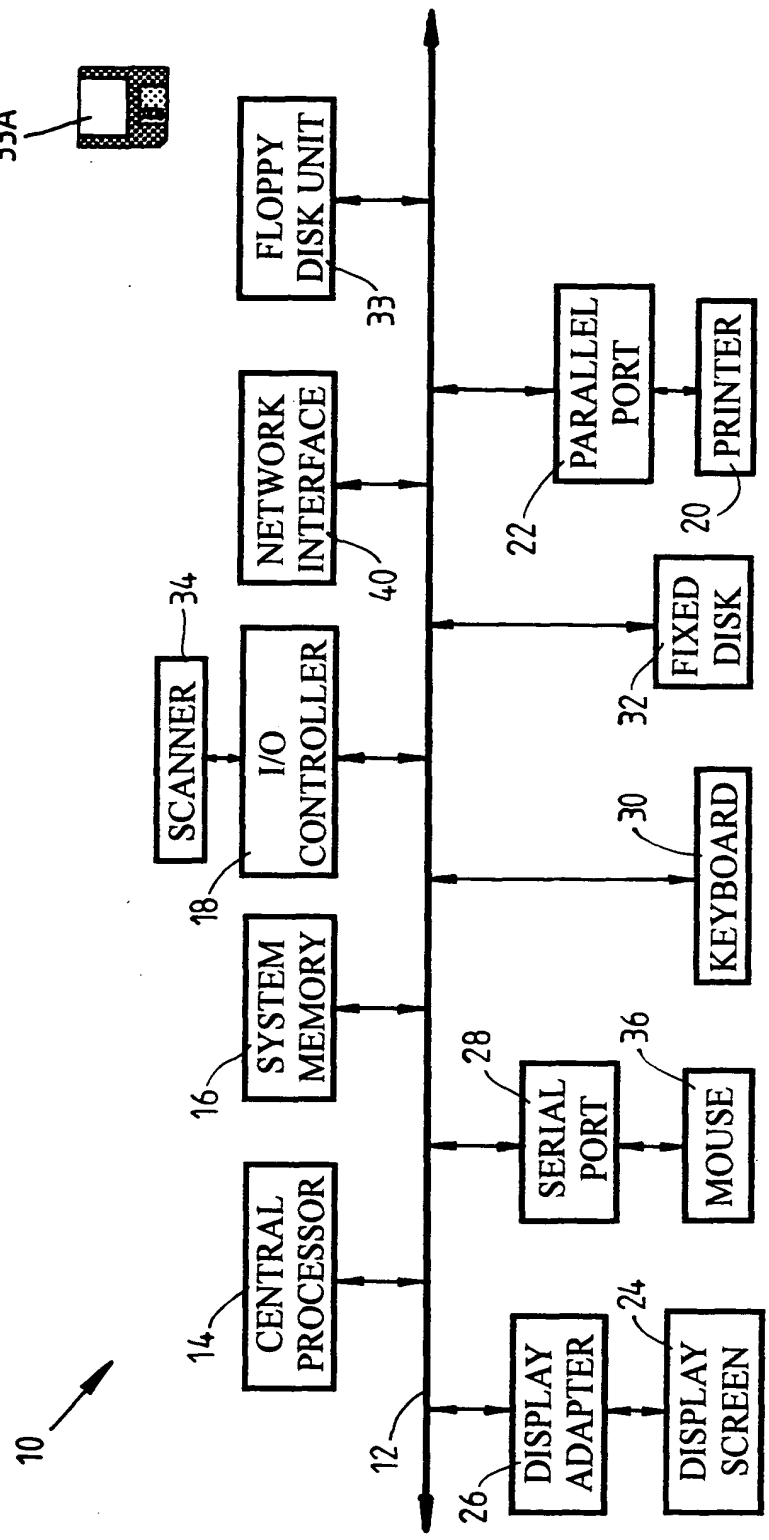

10
12
14
CENTRAL PROCESSOR
16
SYSTEM MEMORY
18
I/O CONTROLLER
20
PRINTER
22
PARALLEL PORT
24
DISPLAY SCREEN
26
DISPLAY ADAPTER
28
SERIAL PORT
30
KEYBOARD
32
FIXED DISK
33
FLOPPY DISK UNIT
33A
34
SCANNER
36
MOUSE
40
NETWORK INTERFACE

Fig. 1

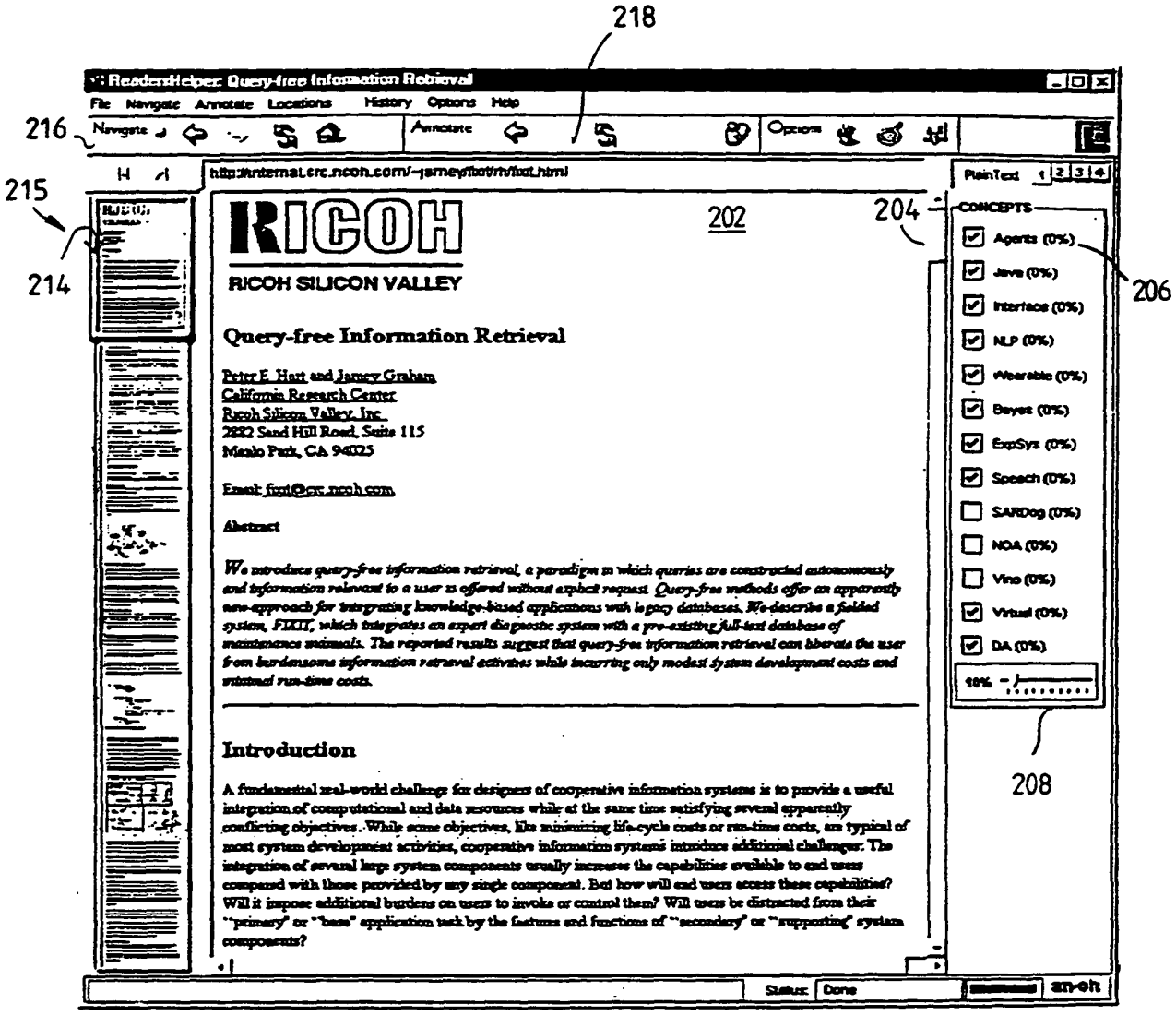

218
216
215
214
204
206
208
202
ReaderHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
PlainText 1 2 3 4
CONCEPTS
Agents (0%)
Java (0%)
Interface (0%)
NLP (0%)
Wearable (0%)
Bayes (0%)
ExpSys (0%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
RICOH
RICOH SILICON VALLEY
Query-free Information Retrieval
Peter E. Hart and Jamey Graham
California Research Center
Ricoh Silicon Valley, Inc.
2882 Sand Hill Road, Suite 115
Menlo Park, CA 94025
Email: hart@crc.ricoh.com
Abstract
We introduce query-free information retrieval, a paradigm in which queries are constructed autonomously and information relevant to a user is offered without explicit request. Query-free methods offer an apparently new approach for integrating knowledge-based applications with legacy databases. We describe a fielded system, FIXIT, which integrates an expert diagnostic system with a pre-existing full-text database of maintenance manuals. The reported results suggest that query-free information retrieval can liberate the user from burdensome information retrieval activities while incurring only modest system development costs and minimal run-time costs.
Introduction
A fundamental real-world challenge for designers of cooperative information systems is to provide a useful integration of computational and data resources while at the same time satisfying several apparently conflicting objectives. While some objectives, like minimizing life-cycle costs or run-time costs, are typical of most system development activities, cooperative information systems introduce additional challenges: The integration of several large system components usually increases the capabilities available to end users compared with those provided by any single component. But how will end users access these capabilities? Will it impose additional burdens on users to invoke or control them? Will users be distracted from their "primary" or "base" application task by the features and functions of "secondary" or "supporting" system components?
Status: Done

Fig. 2A

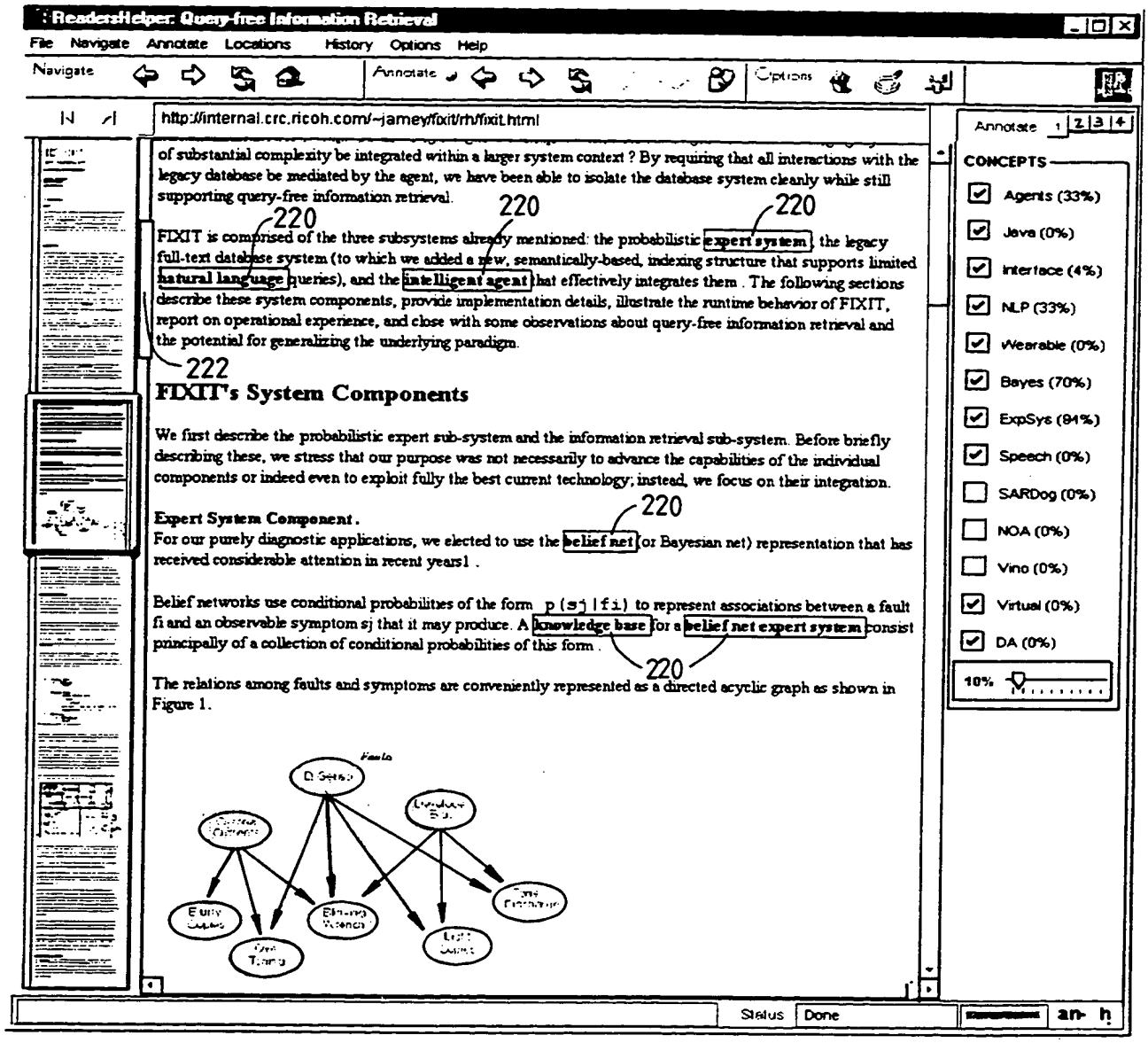

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
http://internal.crc.ricoh.com/~jamey/fixit/rh/fixit.html
of substantial complexity be integrated within a larger system context ? By requiring that all interactions with the legacy database be mediated by the agent, we have been able to isolate the database system cleanly while still supporting query-free information retrieval.
220
220
220
FIXIT is comprised of the three subsystems already mentioned: the probabilistic expert system, the legacy full-text database system (to which we added a new, semantically-based, indexing structure that supports limited natural language queries), and the intelligent agent that effectively integrates them . The following sections describe these system components, provide implementation details, illustrate the runtime behavior of FIXIT, report on operational experience, and close with some observations about query-free information retrieval and the potential for generalizing the underlying paradigm.
222
FIXIT's System Components
We first describe the probabilistic expert sub-system and the information retrieval sub-system. Before briefly describing these, we stress that our purpose was not necessarily to advance the capabilities of the individual components or indeed even to exploit fully the best current technology; instead, we focus on their integration.
220
Expert System Component .
For our purely diagnostic applications, we elected to use the belief net (or Bayesian net) representation that has received considerable attention in recent years1 .
Belief networks use conditional probabilities of the form p(sj|fi) to represent associations between a fault fi and an observable symptom sj that it may produce. A knowledge base for a belief net expert system consist principally of a collection of conditional probabilities of this form .
220
The relations among faults and symptoms are conveniently represented as a directed acyclic graph as shown in Figure 1.
Annotate
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (84%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
10%
Status
Done

Fig. 2B

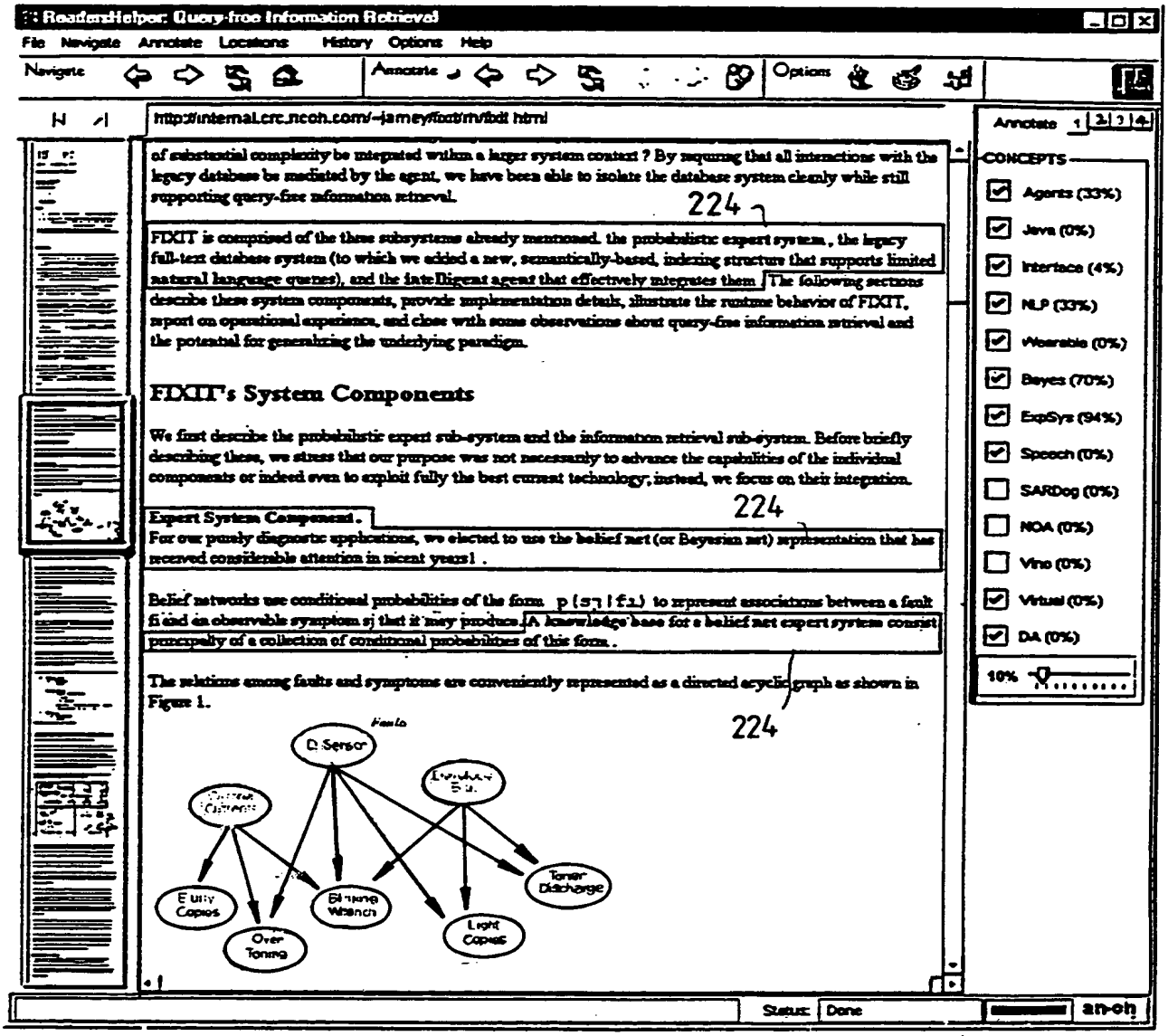

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
http://internal.crc.ricoh.com/~jamey/foo/m/fixit.html
of substantial complexity be integrated within a larger system context ? By requiring that all interactions with the legacy database be mediated by the agent, we have been able to isolate the database system cleanly while still supporting query-free information retrieval.
224
FIXIT is comprised of the three subsystems already mentioned, the probabilistic expert system, the legacy full-text database system (to which we added a new, semantically-based, indexing structure that supports limited natural language queries), and the intelligent agent that effectively integrates them. The following sections describe these system components, provide implementation details, illustrate the runtime behavior of FIXIT, report on operational experience, and close with some observations about query-free information retrieval and the potential for generalizing the underlying paradigm.
FIXIT's System Components
We first describe the probabilistic expert sub-system and the information retrieval sub-system. Before briefly describing these, we stress that our purpose was not necessarily to advance the capabilities of the individual components or indeed even to exploit fully the best current technology; instead, we focus on their integration.
224
Expert System Component.
For our purely diagnostic applications, we elected to use the belief net (or Bayesian net) representation that has received considerable attention in recent years1 .
Belief networks use conditional probabilities of the form p(sj | fi) to represent associations between a fault fi and an observable symptom sj that it may produce. A knowledge base for a belief net expert system consist principally of a collection of conditional probabilities of this form.
The relations among faults and symptoms are conveniently represented as a directed acyclic graph as shown in Figure 1.
224
Faults
D Sensor
Toner Discharge
Light Copies
Over Toning
Annotate
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (94%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
10%
Status: Done

Fig. 2C

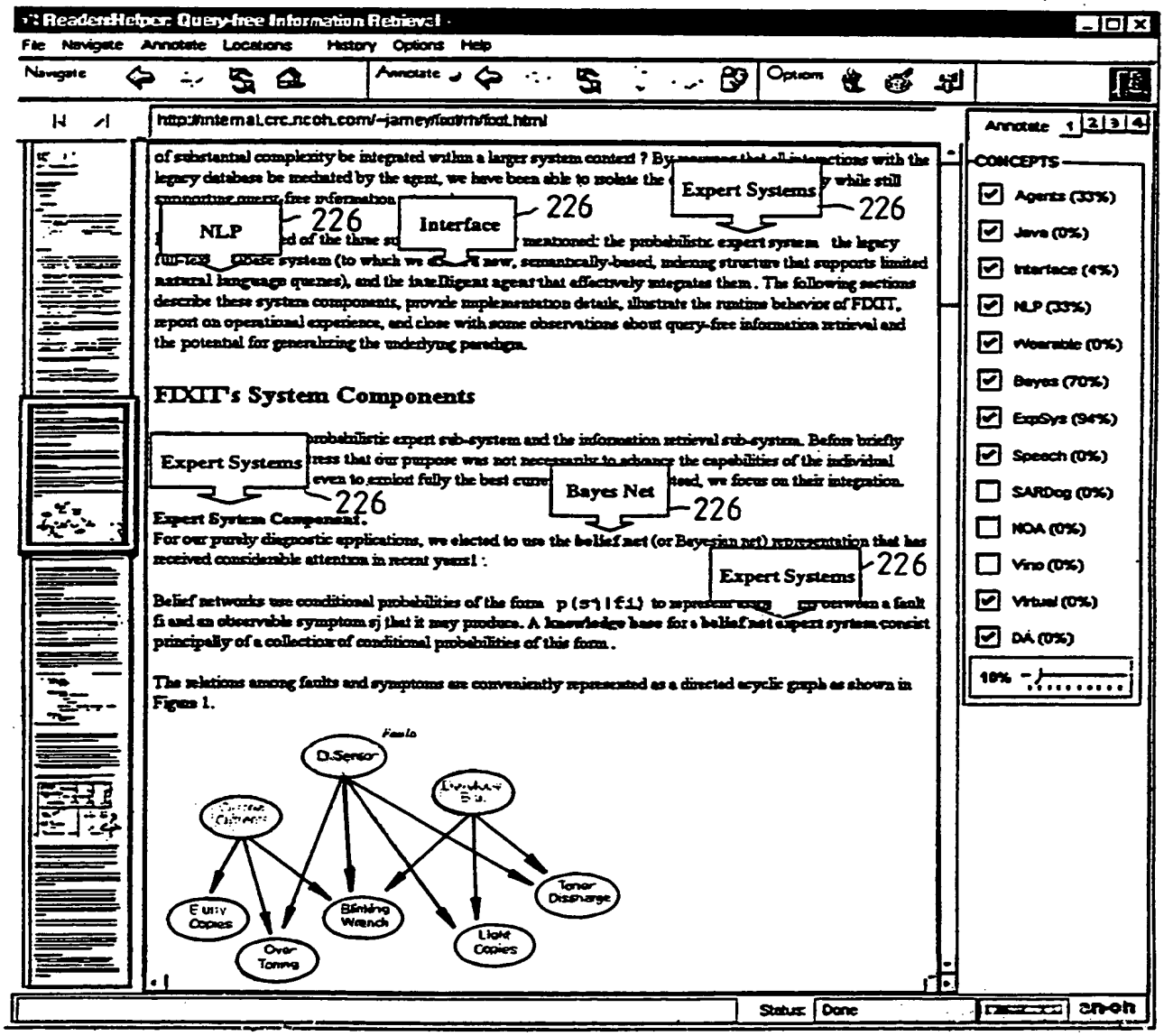

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
Annotate 1 2 3 4
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (94%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
NLP
Interface
Expert Systems
Expert Systems
Bayes Net
Expert Systems
226
FIXIT's System Components
Expert System Component.
Belief networks use conditional probabilities of the form p(sj|fi) to
Status: Done

Fig. 2D

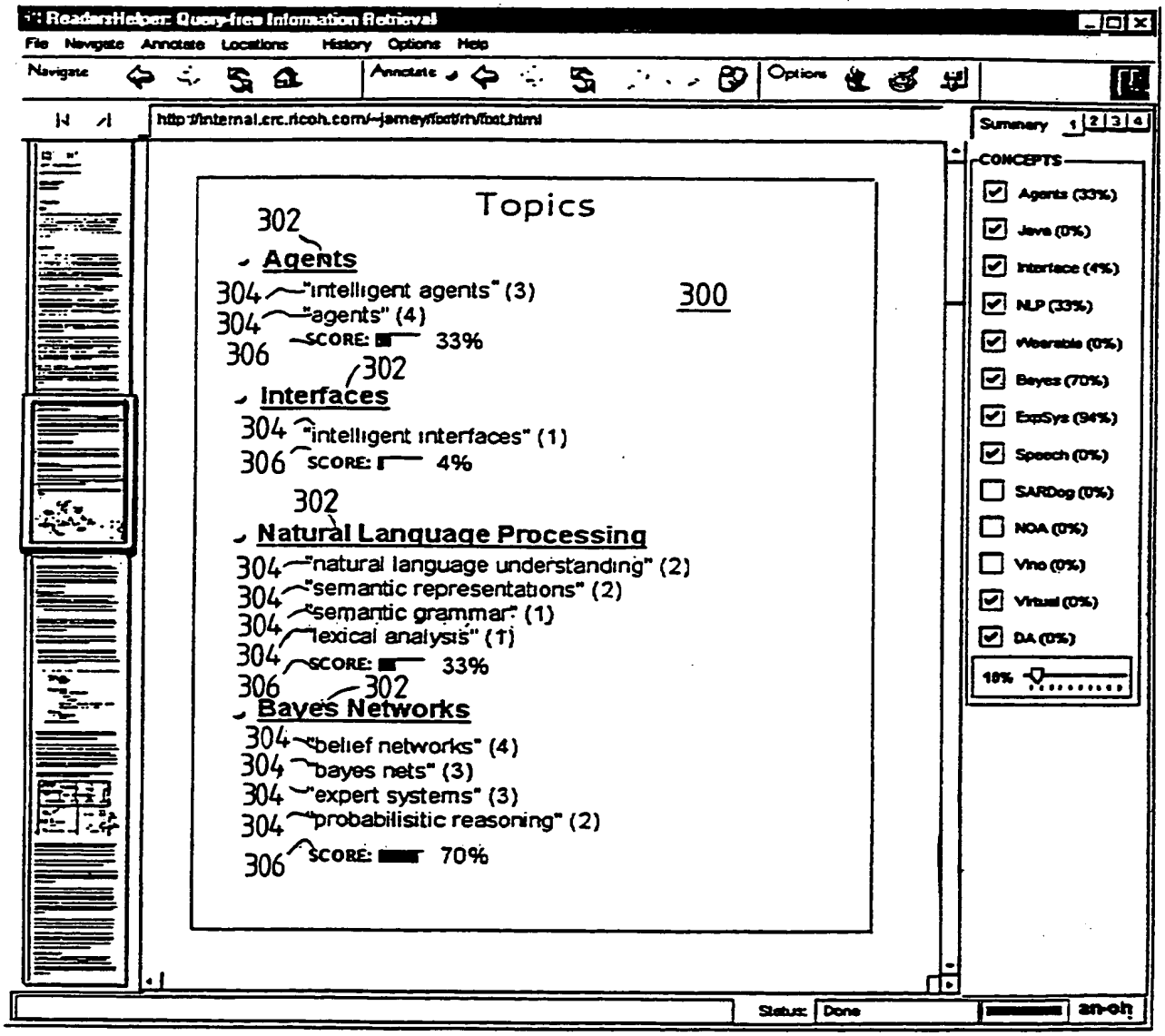

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
http://internal.crc.ricoh.com/~jamey/bot/rh/bot.html
Summary 1 2 3 4
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (94%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
10%
Topics
302
Agents
304 "intelligent agents" (3)
304 "agents" (4)
306 SCORE: 33%
300
302
Interfaces
304 "intelligent interfaces" (1)
306 SCORE: 4%
302
Natural Language Processing
304 "natural language understanding" (2)
304 "semantic representations" (2)
304 "semantic grammar" (1)
304 "lexical analysis" (1)
304 SCORE: 33%
306
302
Bayes Networks
304 "belief networks" (4)
304 "bayes nets" (3)
304 "expert systems" (3)
304 "probabilisitic reasoning" (2)
306 SCORE: 70%
Status: Done

Fig. 3

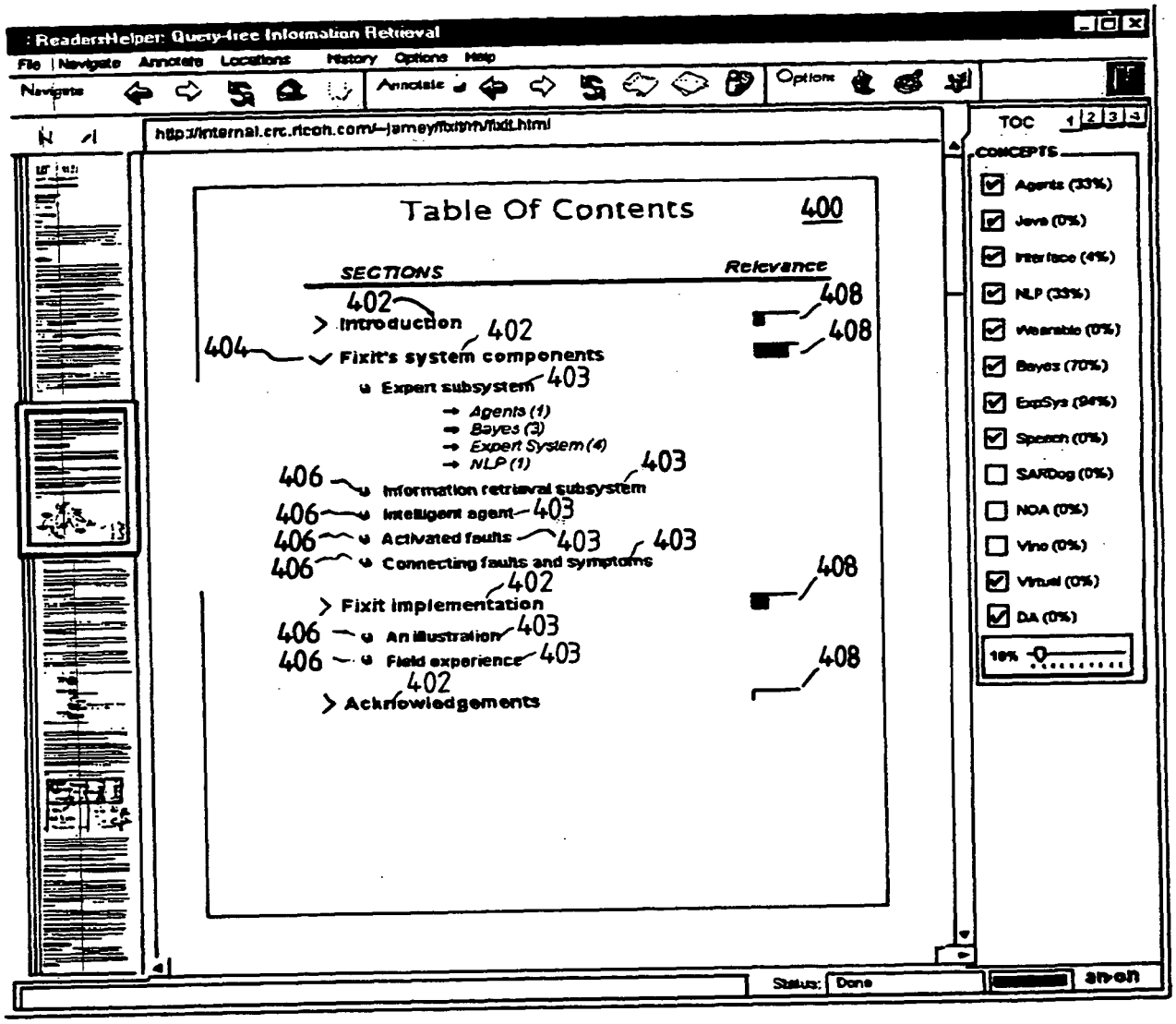

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
http://internal.crc.ricoh.com/~jamey/fixit/fixit.html
Table Of Contents
400
SECTIONS
Relevance
402
Introduction
402
404
Fixit's system components
Expert subsystem
403
Agents (1)
Bayes (3)
Expert System (4)
NLP (1)
406
Information retrieval subsystem
403
406
Intelligent agent
403
406
Activated faults
403
403
406
Connecting faults and symptoms
408
402
Fixit implementation
406
An illustration
403
406
Field experience
403
408
402
Acknowledgements
408
TOC
1 2 3 4
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (94%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
10%
Status:
Done

Fig. 4

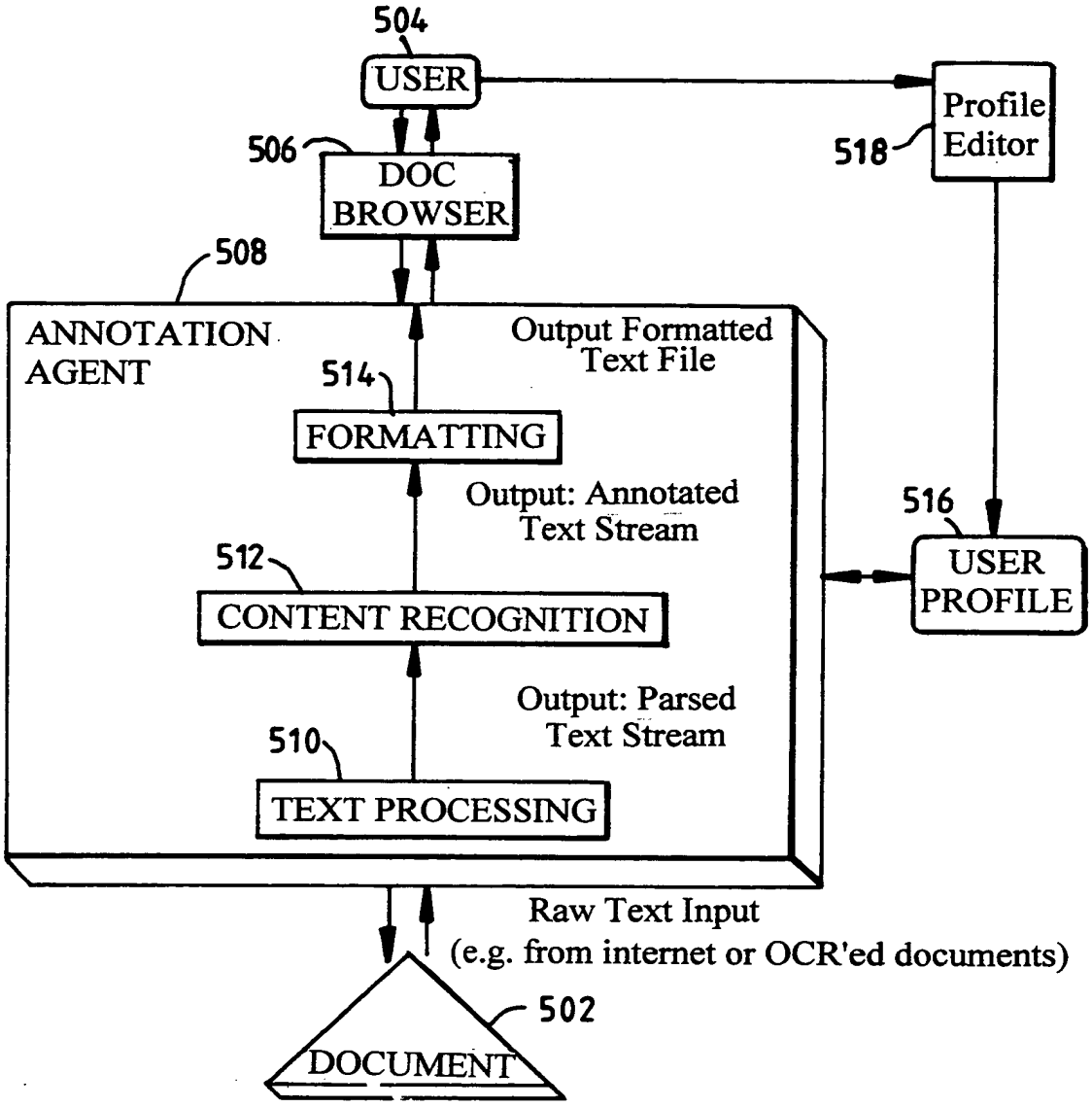

504
USER
Profile Editor
518
506
DOC BROWSER
508
ANNOTATION AGENT
Output Formatted Text File
514
FORMATTING
Output: Annotated Text Stream
516
USER PROFILE
512
CONTENT RECOGNITION
Output: Parsed Text Stream
510
TEXT PROCESSING
Raw Text Input
(e.g. from internet or OCR'ed documents)
502
DOCUMENT

Fig. 5

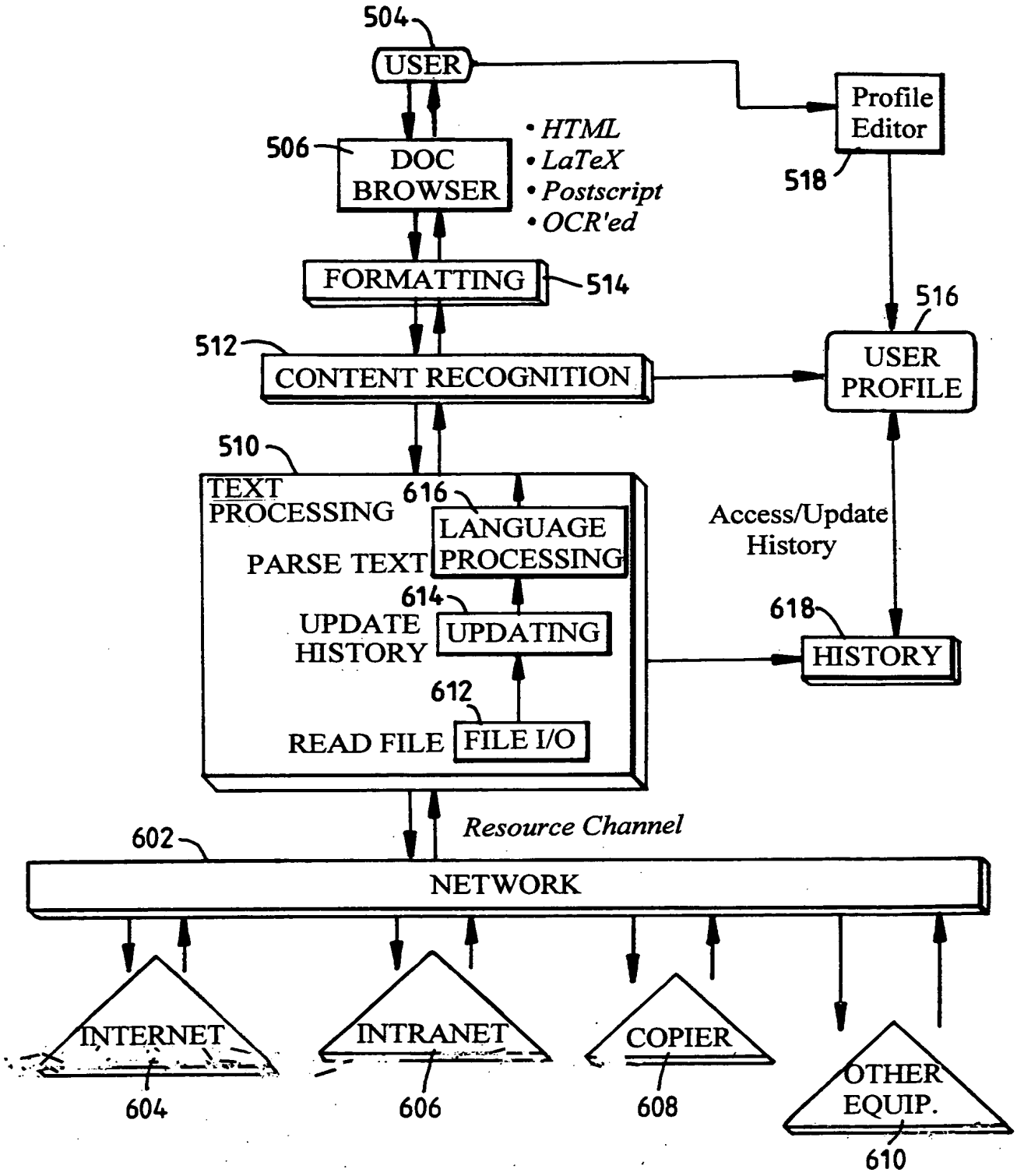

504
USER
Profile Editor
518
506
DOC BROWSER
• HTML
• LaTeX
• Postscript
• OCR'ed
FORMATTING
514
516
512
CONTENT RECOGNITION
USER PROFILE
510
TEXT PROCESSING
616
LANGUAGE PROCESSING
PARSE TEXT
Access/Update History
614
UPDATE HISTORY
UPDATING
618
HISTORY
612
READ FILE
FILE I/O
Resource Channel
602
NETWORK
INTERNET
INTRANET
COPIER
OTHER EQUIP.
604
606
608
610

Fig. 6A

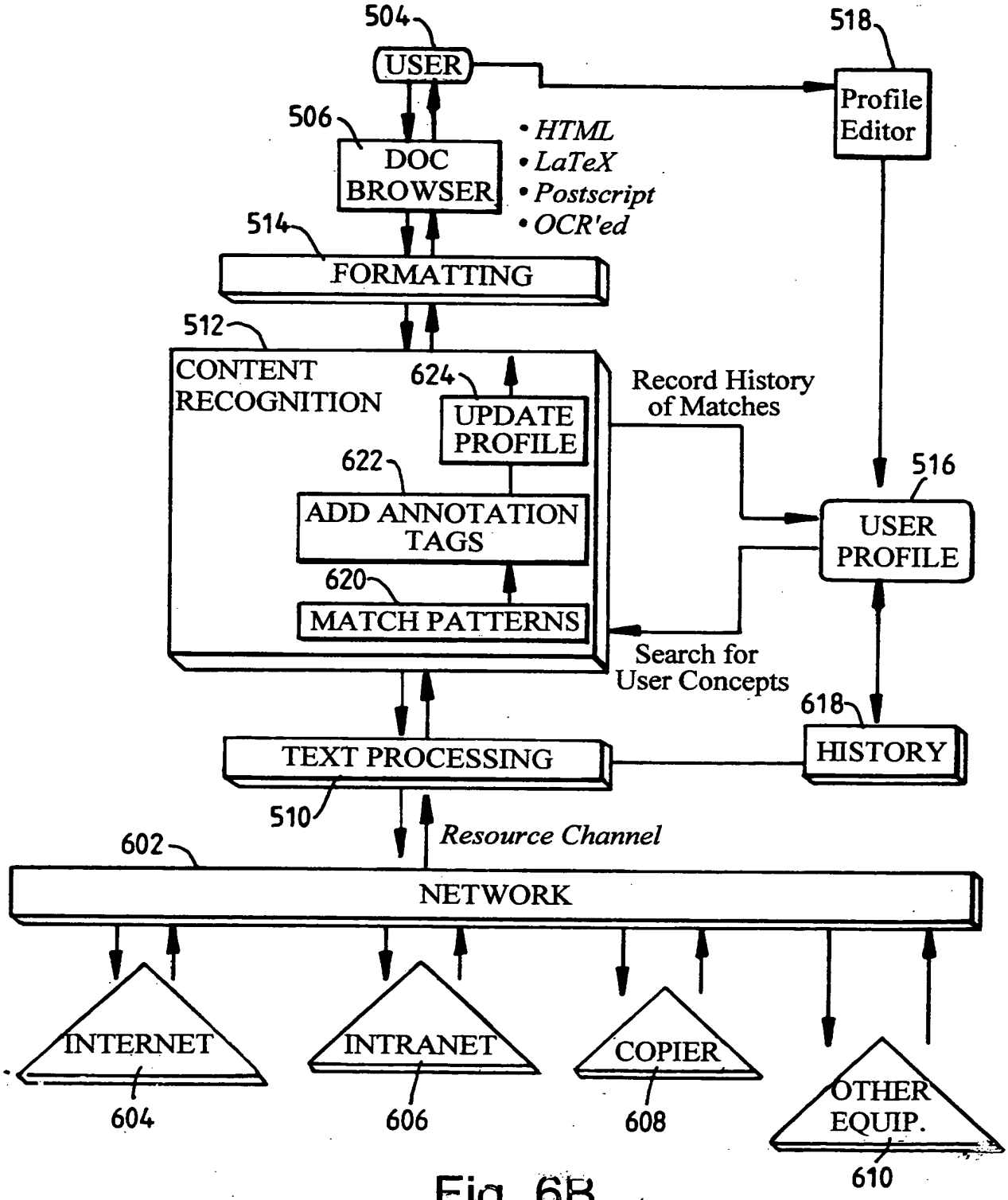

504
USER
518
Profile Editor
506
DOC BROWSER
• HTML
• LaTeX
• Postscript
• OCR'ed
514
FORMATTING
512
CONTENT RECOGNITION
624
UPDATE PROFILE
Record History of Matches
622
ADD ANNOTATION TAGS
516
USER PROFILE
620
MATCH PATTERNS
Search for User Concepts
618
HISTORY
TEXT PROCESSING
510
Resource Channel
602
NETWORK
INTERNET
604
INTRANET
606
COPIER
608
OTHER EQUIP.
610

Fig. 6B

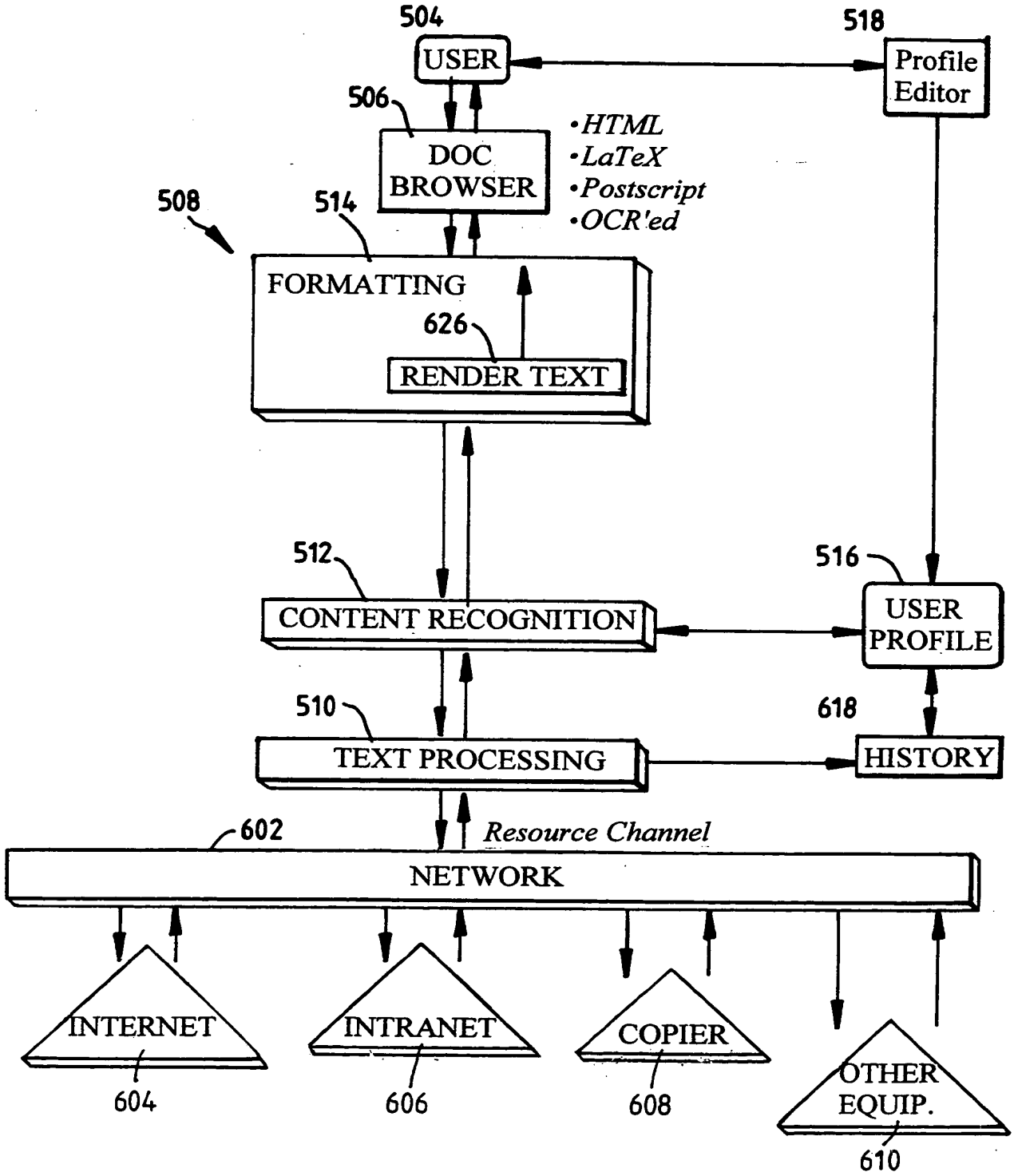

504
USER
518
Profile Editor
506
DOC BROWSER
•HTML
•LaTeX
•Postscript
•OCR'ed
508
514
FORMATTING
626
RENDER TEXT
512
CONTENT RECOGNITION
516
USER PROFILE
510
618
TEXT PROCESSING
HISTORY
602
Resource Channel
NETWORK
INTERNET
604
INTRANET
606
COPIER
608
OTHER EQUIP.
610

Fig. 6C

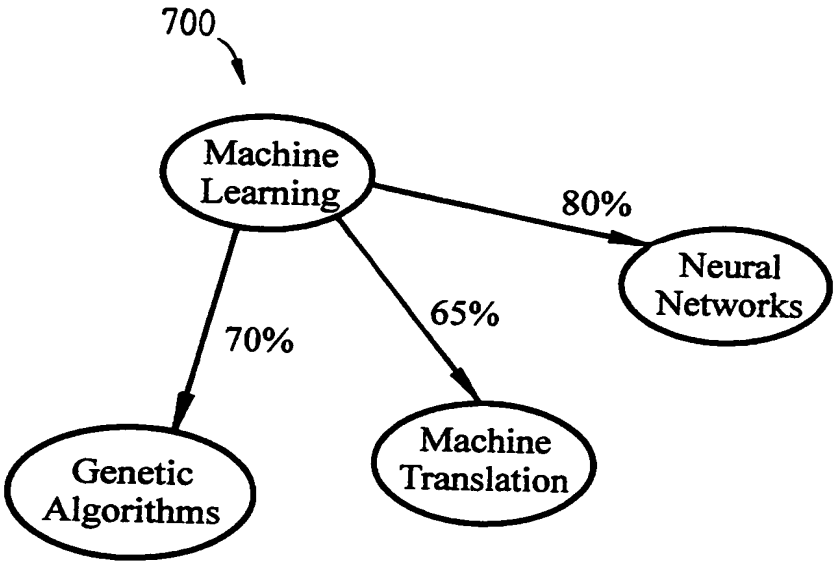

700
Machine Learning
80%
Neural Networks
70%
65%
Genetic Algorithms
Machine Translation

Fig. 7

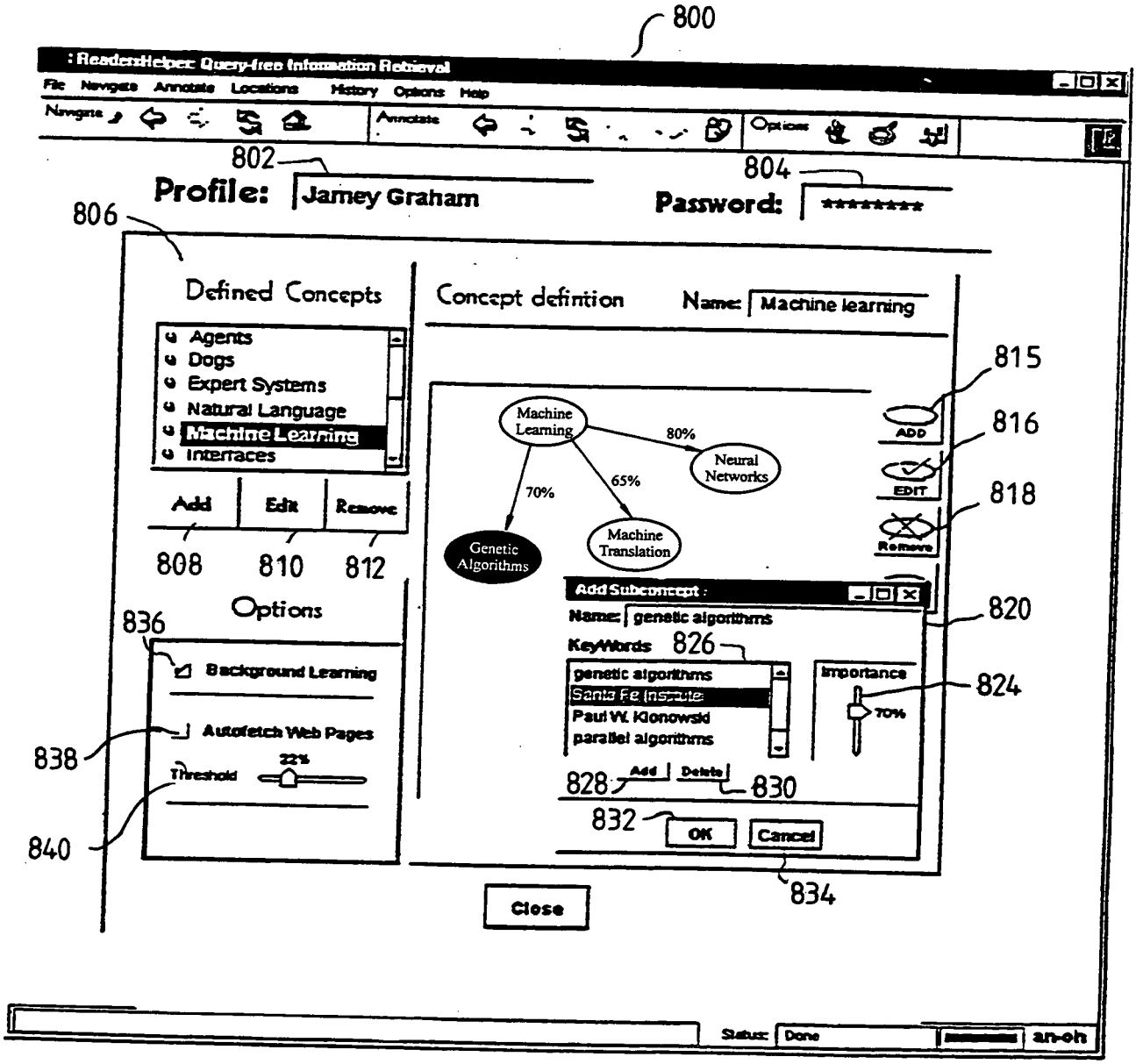

800
ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
802
Profile: Jamey Graham
804
Password: ********
806
Defined Concepts
Agents
Dogs
Expert Systems
Natural Language
Machine Learning
Interfaces
Add
Edit
Remove
808
810
812
Options
836
Background Learning
Autofetch Web Pages
838
Threshold
22%
840
Concept definition
Name: Machine learning
Machine Learning
80%
Neural Networks
70%
65%
Genetic Algorithms
Machine Translation
815
ADD
816
EDIT
818
Remove
Add Subconcept :
Name: genetic algorithms
KeyWords
826
genetic algorithms
Santa Fe Institute
Paul W. Klonowski
parallel algorithms
Importance
824
70%
820
828
Add
Delete
830
832
OK
Cancel
834
Close
Status: Done

Fig. 8

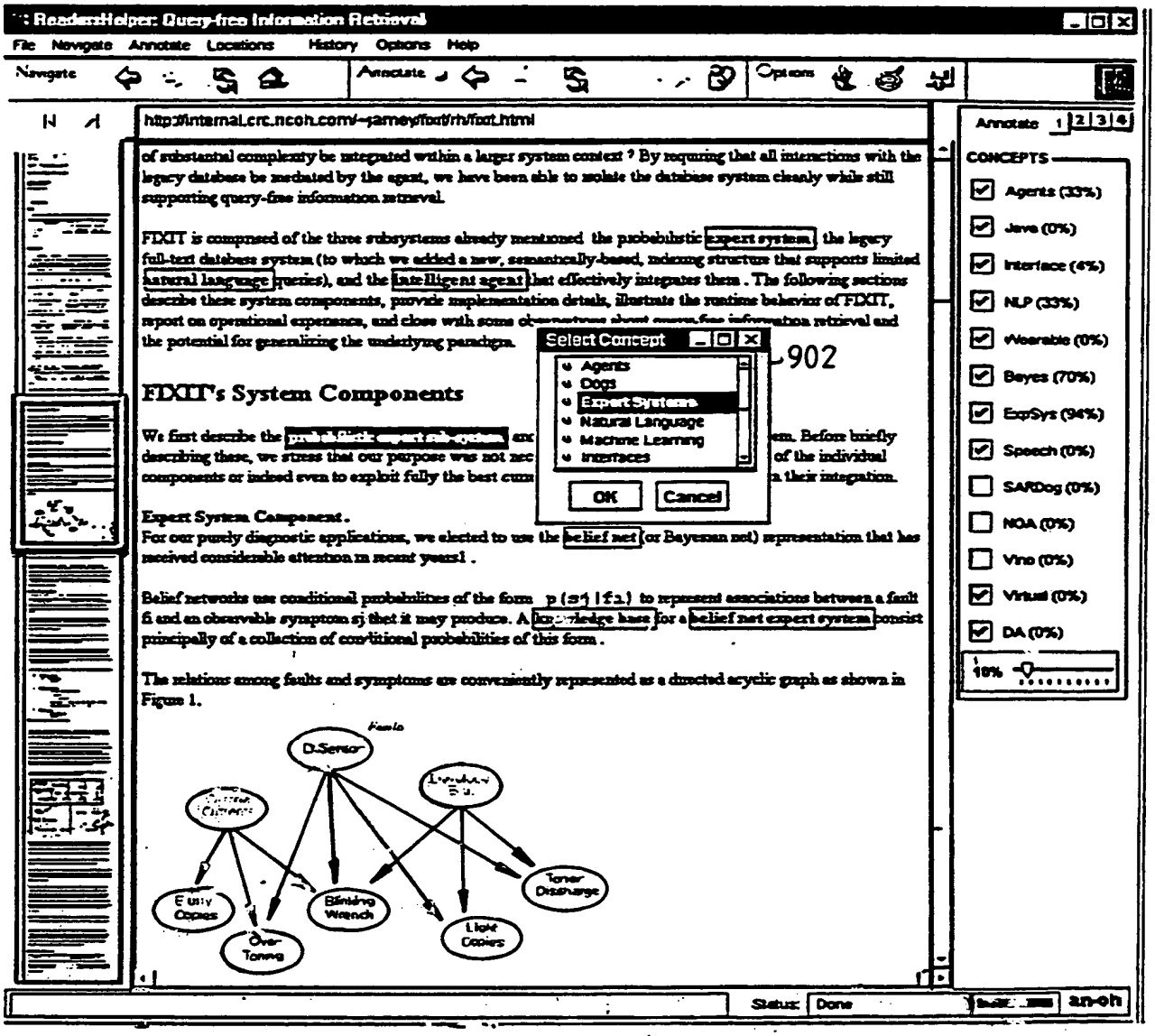

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
http://internal.crc.ricoh.com/~jamey/fixit/rh/fixit.html
of substantial complexity be integrated within a larger system context ? By requiring that all interactions with the legacy database be mediated by the agent, we have been able to isolate the database system cleanly while still supporting query-free information retrieval.
FIXIT is comprised of the three subsystems already mentioned the probabilistic expert system, the legacy full-text database system (to which we added a new, semantically-based, indexing structure that supports limited natural language queries), and the intelligent agent that effectively integrates them . The following sections describe these system components, provide implementation details, illustrate the runtime behavior of FIXIT, report on operational experience, and close with some observations about query-free information retrieval and the potential for generalizing the underlying paradigm.
Select Concept
Agents
Dogs
Expert Systems
Natural Language
Machine Learning
Interfaces
OK
Cancel
902
FIXIT's System Components
We first describe the
em. Before briefly describing these, we stress that our purpose was not nec
of the individual components or indeed even to exploit fully the best curr
their integration.
Expert System Component.
For our purely diagnostic applications, we elected to use the belief net (or Bayesian net) representation that has received considerable attention in recent years1 .
Belief networks use conditional probabilities of the form p(sj|fi) to represent associations between a fault fi and an observable symptom sj that it may produce. A knowledge base for a belief net expert system consist principally of a collection of conditional probabilities of this form .
The relations among faults and symptoms are conveniently represented as a directed acyclic graph as shown in Figure 1.
D-Sensor
Toner Discharge
Blinking Wrench
Light Copies
Over Toning
Annotate 1 2 3 4
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (94%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
Status: Done

Fig. 9A

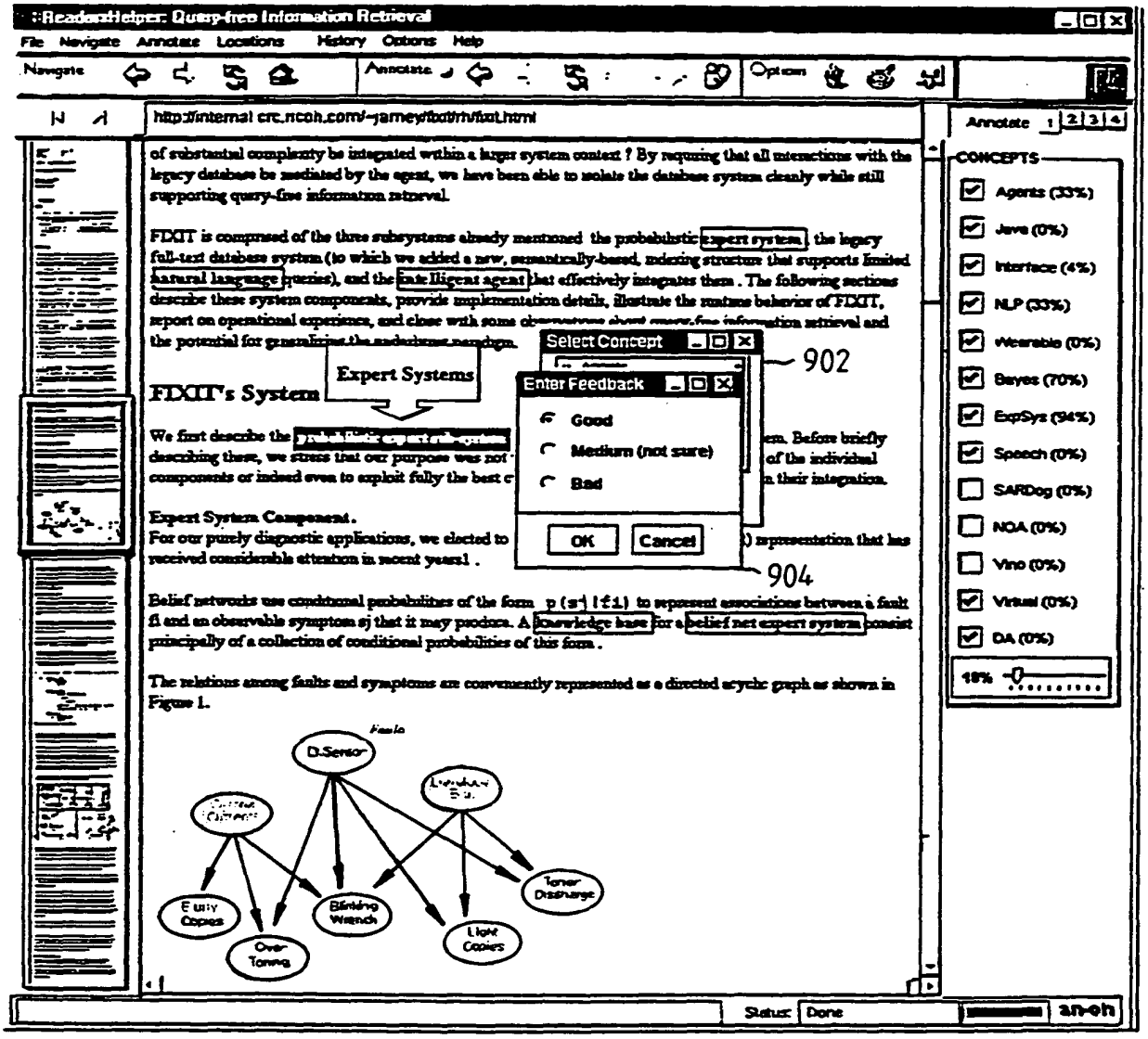
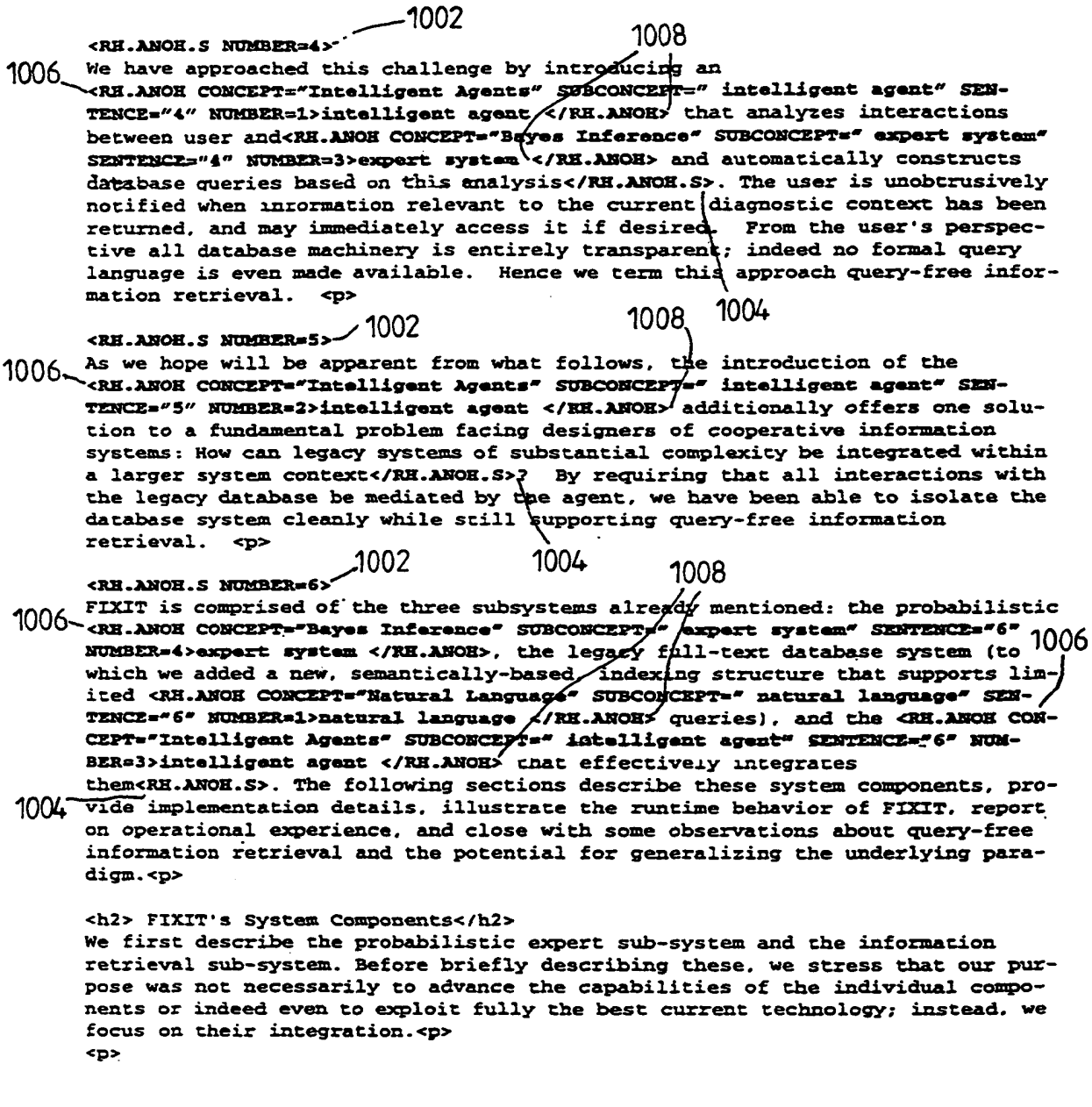

ReadersHelper: Query-free Information Retrieval
File Navigate Annotate Locations History Options Help
Navigate
Annotate
Options
http://internal.crc.ncoh.com/~jamey/foo/m/foo1.html
of substantial complexity be integrated within a larger system context ? By requiring that all interactions with the legacy database be mediated by the agent, we have been able to isolate the database system cleanly while still supporting query-free information retrieval.
FIXIT is composed of the three subsystems already mentioned the probabilistic expert system, the legacy full-text database system (to which we added a new, semantically-based, indexing structure that supports limited natural language queries), and the intelligent agent that effectively integrates them . The following sections describe these system components, provide implementation details, illustrate the runtime behavior of FIXIT, report on operational experience, and close with some observations about query-free information retrieval and the potential for generalizing the underlying paradigm.
Expert Systems
FIXIT's System
Select Concept
902
Enter Feedback
Good
Medium (not sure)
Bad
OK
Cancel
904
We first describe the
em. Before briefly describing these, we stress that our purpose was not
of the individual components or indeed even to exploit fully the best c
n their integration.
Expert System Component.
For our purely diagnostic applications, we elected to
) representation that has received considerable attention in recent years1 .
Belief networks use conditional probabilities of the form p(sj|fi) to represent associations between a fault fi and an observable symptom sj that it may produce. A knowledge base for a belief net expert system consist principally of a collection of conditional probabilities of this form .
The relations among faults and symptoms are conveniently represented as a directed acyclic graph as shown in Figure 1.
Faults
D.Sensor
Toner Discharge
Blinking Wrench
Light Copies
Over Toning
Annotate 1 2 3 4
CONCEPTS
Agents (33%)
Java (0%)
Interface (4%)
NLP (33%)
Wearable (0%)
Bayes (70%)
ExpSys (94%)
Speech (0%)
SARDog (0%)
NOA (0%)
Vino (0%)
Virtual (0%)
DA (0%)
18%
Status: Done

Fig. 9B



<RH.ANOH.S NUMBER=4>
We have approached this challenge by introducing an
<RH.ANOH CONCEPT="Intelligent Agents" SUBCONCEPT=" intelligent agent" SENTENCE="4" NUMBER=1>intelligent agent </RH.ANOH> that analyzes interactions between user and<RH.ANOH CONCEPT="Bayes Inference" SUBCONCEPT=" expert system" SENTENCE="4" NUMBER=3>expert system </RH.ANOH> and automatically constructs database queries based on this analysis</RH.ANOH.S>. The user is unobtrusively notified when information relevant to the current diagnostic context has been returned, and may immediately access it if desired. From the user's perspective all database machinery is entirely transparent; indeed no formal query language is even made available. Hence we term this approach query-free information retrieval. <p>

<RH.ANOH.S NUMBER=5>
As we hope will be apparent from what follows, the introduction of the
<RH.ANOH CONCEPT="Intelligent Agents" SUBCONCEPT=" intelligent agent" SENTENCE="5" NUMBER=2>intelligent agent </RH.ANOH> additionally offers one solution to a fundamental problem facing designers of cooperative information systems: How can legacy systems of substantial complexity be integrated within a larger system context</RH.ANOH.S>? By requiring that all interactions with the legacy database be mediated by the agent, we have been able to isolate the database system cleanly while still supporting query-free information retrieval. <p>

<RH.ANOH.S NUMBER=6>
FIXIT is comprised of the three subsystems already mentioned: the probabilistic
<RH.ANOH CONCEPT="Bayes Inference" SUBCONCEPT=" expert system" SENTENCE="6" NUMBER=4>expert system </RH.ANOH>, the legacy full-text database system (to which we added a new, semantically-based, indexing structure that supports limited <RH.ANOH CONCEPT="Natural Language" SUBCONCEPT=" natural language" SENTENCE="6" NUMBER=1>natural language </RH.ANOH> queries), and the <RH.ANOH CONCEPT="Intelligent Agents" SUBCONCEPT=" intelligent agent" SENTENCE="6" NUMBER=3>intelligent agent </RH.ANOH> that effectively integrates them<RH.ANOH.S>. The following sections describe these system components, provide implementation details, illustrate the runtime behavior of FIXIT, report on operational experience, and close with some observations about query-free information retrieval and the potential for generalizing the underlying paradigm.<p>

<h2> FIXIT's System Components</h2>
We first describe the probabilistic expert sub-system and the information retrieval sub-system. Before briefly describing these, we stress that our purpose was not necessarily to advance the capabilities of the individual components or indeed even to exploit fully the best current technology; instead, we focus on their integration.<p>
<p>

Fig. 10

The Patent Office

21

Application No:	GB 9827135.6	**Examiner:**	Mike Davis
Claims searched:	1-11,14-25,28-31	**Date of search:**	27 January 1999

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.Q): G4A (AUXX, AUDB)

Int Cl (Ed.6): G06F

Other:

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X	EP 0802492 A1	(IBM) eg pages 2-9 and Figs.11-15	1,2,14,28, 30 at least
X	EP 0762297 A2	(SUN) eg pages 2-5, and page 10 line 52 to page 11 line 39	"
X	EP 0378848 A2	(IBM) eg abstract and page 4 lines 18-20	"
X	US 5404295	(KATZ ET AL) eg columns 1-4, and column 12 line 50 to column 13 line 24	"

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.